Evaluating the terminology requirements to support multi-disciplinary diabetes care

Philip J. B. Brown, MRCGP, Colin Price FRCS,
*Peter H. Sonksen MD, FRCP.
National Health Service Centre for Coding and Classification,
Loughborough, UK
*Department of Endocrinology, Diabetes & Metabolic Medicine,
St Thomas's Hospital, London, UK

Diabetes mellitus is a multi-system disease requiring lifetime multi-disciplinary care, which has proven individual and economic benefits. The delivery of service involves co-operation and communication between patient, carer and health care professionals, and systematic auditing of processes and outcomes. Sustained improvement necessitates regular data acquisition, aggregation and analysis. The terminology requirements to support patient-centred records and identified datasets are examined, and differences in purpose and scope highlighted. The many stakeholders involved in diabetes care have their own sublanguages and terminology requirements which need harmonising around a common core. The problems and solutions of accommodating these needs are explored in relation to the Read Thesaurus.

INTRODUCTION

The care of patients with diabetes mellitus provides an excellent test bed for the evaluation of clinical terminologies. The disorder is multi-system and requires long term care involving multi-professional input and co-operation. It involves a wide variety of processes including screening, monitoring, interventions and education of both the patient and any carer involved. The attainment of permanent, population-wide improvement requires the recording and use of data to complete the feedback cycle between outcomes and interventions (medical and behaviour measures) by all stakeholders (patients, carers and health care professionals).[1] Clinical vocabularies and datasets both have important roles in the recording and analysis of data on diabetes but differ in purpose, scope and content.

BACKGROUND

Diabetes mellitus is important both in personal and economic terms.[2] The value of prevention, good control and effective intervention in the reduction of complications has been demonstrated.[3] The importance of earlier diagnosis and treatment has been recently highlighted by the American Diabetic Association recommending that everyone older than 45 should be screened for diabetes.[4]

St Vincent Declaration

The recognition of the need to improve diabetes care across Europe was identified during the Eurodiabeta AIM project in 1989. This led to a joint initiative between the European branch of the World Health Organisation and the International Diabetes Federation culminating in the publication of the *St Vincent Declaration*.[5] This identifies specific targets relating to a reduction in mortality, blindness, end-stage renal failure, limb amputation, ischaemic heart disease and achieving pregnancy outcomes approaching those of non-diabetic women. A further recommendation of the Declaration was to realise the potential of information technology in attainment of these goals. Within the United Kingdom the methodology for the implementation of the recommendations was devised by the Joint Task Force for Diabetes, cosponsored by the Department of Health (DOH) and the British Diabetes Association. One of the products of this work was the development of the *UK diabetes dataset*.[6] On a European platform a number of collaborative projects have been stimulated which included the development of the *DiabCare dataset*.[7] These datasets include greater detail than is available from the International Statistical Classification of Diseases and Related Health Problems - Tenth revision (ICD10),[8] although only partly satisfy the information requirements to support the full Electronic Patient Record (EPR).

The Read Thesaurus

The Read Thesaurus was developed during the Clinical Terms Project as a user-led terminology covering all aspects of health care.[9] The 55 Specialty Working Groups included the following domains:

- Diabetes
- Endocrinology

- Dietetics
- General Practice
- Chiropody and Podiatry
- Nursing professions
- Ophthalmology
- Vascular surgery
- Neurology
- Obstetrics
- Orthopaedics
- Renal medicine
- Child health

Following the initial collection exercise, the concepts were integrated and continue to be refined and expanded in response to user feedback.

During this process significant experience was gained in the recognition of alternative representation of concepts, and in the reconciliation of these views both politically and practically within the thesaurus.[10] The Read Thesaurus thus represents a substantial resource to support the terminology requirements of the different professionals involved in diabetes care.

PATIENT-CENTRED TERMINOLOGY

Coding schemes are needed to support different levels of information requirements. Formal classifications such as ICD10 are designed for the collection of aggregate data for national and international comparisons. In contrast, clinical vocabularies contain the detail needed to support direct patient care.[11] A number of ideal features of controlled clinical vocabularies have been described to support the EPR and meet the requirements of individual patient care.[12,13] In order to maximise its potential to facilitate multi-disciplinary care a terminology needs to have:

- Shared ownership
- Professional support
- Completeness (in scope and detail)
- Filterability (providing different views)

The Clinical Terms Project undertook the development of a user-led terminology to meet these identified requirements. It was a large undertaking, and required reconciliation between different professional groups and mechanisms to support different views of the common thesaurus. During the development of the Read Thesaurus it was acknowledged that each group had their own *specialist terminology* requirements that had to be both harmonised into a common thesaurus, whilst retaining information to allow construction of

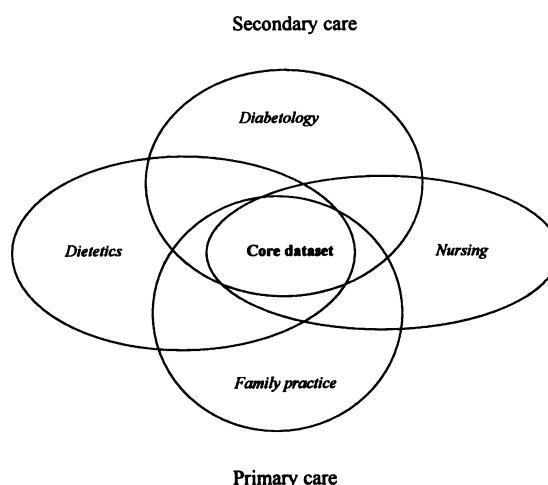individual views by filtering mechanisms including *subsets* and *natural synonyms*.

Secondary care

Diabetology

Dietetics   Core dataset   Nursing

Family practice

Primary care

Figure 1 - Sublanguages involved in diabetes care

Sublanguages

The care of diabetes involves a number of stakeholders including the patient, carer and a variety of health care professionals. The success of education, screening, prevention, control, interventions, and ultimately outcomes, is dependant upon how each of these components of care are harmonised. Each of these parties have their own requirement for terminology to support and express their role which can be described as sublanguages.[14] These sublanguages have their own specific domains of terminology and also contain some elements that are of common significance to others. Figure 1 illustrates a small number of the different professional sublanguages involved in diabetes care which would also include ophthalmology, chiropody, cardiology, neurology, vascular surgery etc.

Specialist terminology

Some professional groups require large numbers of specialist concepts without which it would be difficult to capture their records. For example, within the Read Thesaurus the dietetics profession had to develop detailed hierarchies of foods and drinks, nutrients, food additives, dietary advice, diets and dietary intake to capture their most basic activity (table 1).

The incorporation of such specialist areas into a thesaurus is a large task and frequently requires the development of new areas of terminology where no established vocabularies are available as source material

Table 1 - Numbers of unique concepts required to support dietetic terminology in the Read Thesaurus

Hierarchy	Number of concepts
Foods and drinks	1899
Nutrients	171
Food additives	61
Dietary advice	129
Diets	227
Dietary intake	217

Subsets

The large size of a terminology accommodating the required specialist detail ideally should include a mechanism for filtering concepts.[12] This feature allows different specialist subsets to be used from a shared terminology by improving retrieval times, duration of term searching and shorter relevant picking lists. Different varieties of subsets may be required and their generation needs to take account of the dynamic nature of the thesaurus.[15]

Natural synonyms

The recognition of the existence of sublanguages is important, so that concepts that might be expressed by using synonymous terms are identified. Just as importantly, the sublanguages of different professional groups might use the same term string to express disparate concepts. For example the term "Manipulation of joint" might be used by an orthopaedic surgeon to mean the physical manipulation of a joint under an anaesthetic, whereas a physiotherapist may use it to refer to the regimen of manual therapy to a joint. This problem is well recognised and, for example, within the Unified Medical Language System (UMLS),[16] such strings with multiple ambiguous meanings are dealt with by having a unique string identifier (SUI) linked to more than one concept. The Read Thesaurus copes with this potential problem similarly by allowing a term (represented by a term identifier) to be a *natural synonym* and connected to more than one concept whose meaning is unambiguously expressed by the preferred term (figure 2).
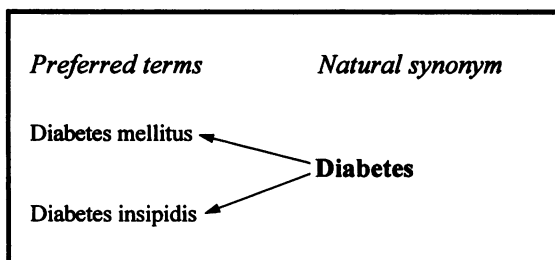

Figure 2 - Natural synonymy of concepts

DATASET TERMINOLOGY

Datasets enable structured collection of multiple variables which represent the key features of care, identified as important markers for the appraisal of efficacy and quality. In order to achieve widespread implementation and thus maximise its potential a dataset needs to have:

- Shared ownership
- Professional support
- Clear objectives
- Development by consensus

The importance of shared ownership and professional support are common requirements with those that have been identified for controlled clinical vocabularies. Indeed a dataset should ideally be created by consensus distillation from such a source. Those concepts of common interest, and overlap between the various sublanguages, form the basis of the creation of a common *core dataset* (figure 1). It is instructive to note that the greater the number of sublanguages involved the smaller the potential overlap. This leads to the tension of providing a core that may be common to all but satisfies none, or a comprehensive set that may be too large to implement.

The St Vincent Declaration gives clear targets and forms a firm framework of objectives around which datasets can be built. The concepts within these datasets can be divided into broad concept types which contain similar characteristics (table 2).

Table 2 - Examples of concept types existing within the diabetes datasets

Concept type	Examples
Disorder	Proliferative retinopathy
Qualitative observation	Absent foot pulses, Injection sites abnormal
Quantitative observation	Body mass index, Albumin excretion rate
Procedure	Retinal photography
Context	Myocardial infarction last year
Demographic	Gender
Record	Date of record

An examination of ICD10, the UK Diabetes, and DiabCare datasets reveals differences in both total number and types of concepts included (table 3). The total number of concepts contained in each set is an empirical measure of the detail, and the percentages of different types of concept a guide to the function for which each were designed.

The concepts included in the analysis of ICD10 are those where a unique code or unique combination (dagger-asterisk) exists. The limitation of recording mortality in diabetes using single coding has been recognised as sub-optimal as there are many other categories within ICD10, such as cardiovascular diseases, which are important and can only be overcome by routine international multiple cause encoding.[17]

Table 3 - Terminological concept types of schemes relating to diabetes (percentages in parentheses)

Concept type	ICD10	UK dataset	DiabCare
Disorder	60 (94)	13 (15)	16 (15)
Qualitative obs	1 (1.5)	17 (20)	11 (11)
Quantitative obs	-	18 (21)	18 (17)
Procedure	2 (3)	18 (21)	14 (13)
Context	1 (1.5)	3 (4)	9 (9)
Demographic	-	3 (4)	4 (4)
Record	-	13 (15)	33 (31)
Total	**64**	**85** [a]	**105** [a]

[a] *The total number of concepts for these datasets are reduced as some are duplicated for right and left*

The more clinically-oriented datasets, as expected, contain greater detail relating to both observations and procedures as compared to ICD10 which is designed purely for categorising indices of mortality and morbidity. These reflect a need to represent a mixture of outcome measures (e.g. the incidence of a disorder), and qualitative and quantitative observations important in the process of management, risk assessment and monitoring of severity and control.

An additional feature of the UK dataset is the allocation of a score to identify whether a concept is present, absent or not known. (For example the entry of 1, 0 or -1 in field 24 indicates whether retinal photography was performed, not performed or not recorded). This explicit modification of the use of a dataset is often overlooked and frequently an assumption is made that absence of an entry implies absence of the finding or action. This illustrates the importance of the relationship between terminologies, datasets, and the context of their intended use within a record structure model.

DISCUSSION

The quality assurance of diabetes care is dependant upon not only an adequate terminology but organisational issues concerning treatment protocols, contracts for care and the availability of adequate clinical information systems.[18] The terminology to support individualised patient diabetes care, and to capture the necessary concepts for each professional involved needs to be detailed and have mechanisms for reconciling these different views.

An important requirement in the development, and influence on success, of both controlled clinical vocabularies and datasets is shared professional ownership and support. In order to avoid bias in construction from any particular viewpoint the creation of the Read Thesaurus was ambitious in attempting to produce a common thesaurus with simultaneous involvement of a large number of representative clinical professional working groups.[10] The success of the project was dependant upon the goodwill of the groups and sometimes upon intensive negotiation. In order to achieve reconciliation between differing professional detail requirements and term uses, a number of mechanisms within the thesaurus such as multiple classification, qualifying detail using a template mechanism,[9] natural synonymy and subsets were employed. These solutions to the problem of achieving completeness and filterability (providing different views) will hopefully be borne out by further operational testing.

The identification of manageable and agreed datasets is a practical way forward in collecting consistent, complete and comparable data. These ideally should not be created in isolation of terminologies supporting the EPR, but should be a distillation by consensus to meet identified objectives. The development of the UK Dataset, and more recently the UK DiabCare dataset,[19] harnessed representative professional members, many of whom were involved in the creation of the Read Thesaurus or were current users. The availability of a common source of identified unique concepts, coupled with the clear objectives set out within the St. Vincent Declaration, has proved a useful solution in identifying core datasets.

The expression of key aspects of care must be integrated with important contextual information such as to whom, and by whom, a procedure was performed; and the recording of negation when, for example, an examination or event did not occur. The unambiguous recording, retrieval and transfer of this information is also dependant upon an agreed record structure.[20]

CONCLUSION

The evaluation and improvement of diabetes care requires information at different levels of detail.

The patient, carer and varied professional groups involved each have their own sublanguages which are vital to record their practice. These are by necessity detailed, and need to be reconciled with each other. Domains of concepts are shared between these groups, and form the basis of collectively endorsed datasets for the exchange of information.

Acknowledgements

References

1. Henrichs HR. Piwernetz K. Sonksen PH. Jahrig D. Reach G. Fischer U. The feedback between monitoring and improvement of quality of diabetes. Diabete et Metabolisme 1993, 19: 70-73.
2. Gerard K, Donaldson C, Maynard A. The cost of diabetes. Diabetic medicine, 1989; 6: 164-170.
3. The DCCT Research Group. The effect of intensive treatment of diabetes on the development and progression of long term complications in insulin dependant diabetes mellitus. New England Journal of Medicine, 1993, 329: 977-986.
4. Macready N. American Diabetes Association calls for testing all those over 45. BMJ, 1997; 315: 11.
5. World Health Organisation (Europe) and International Diabetes Federation (Europe). Diabetes care and research in Europe: The St Vincent Declaration. Diabetic Medicine 1990; 7: 360-360.
6. Vaughan NJA, Home PD, for the Diabetes Audit Working Group of the Research Unit of the Royal College of Physicians and the British Diabetic Association. The UK Diabetes Dataset: A Standard for Information Exchange. Diabetic Medicine, 1995, 12: 717-722.
7. Piwernetz, K, Home PD, Snorgaard O, Antsiferov M, Staehr-Johansen K, Krans M, for the DiabCare Monitoring Group of the St Vincent Declaration Steering Committee. Monitoring the targets of the St Vincent Declaration and the implementation of quality management in diabetes care: The DiabCare initiative. Diabetic Medicine ,1993; 10: 371-377.
8. WHO: International Statistical Classification of Diseases and Related Health Problems. 10th Revision. Geneva: WHO, 1992.
9. O'Neil MJ, Payne C, Read JD. Read Codes Version 3: A User Led Terminology. Meth Inform Med, 1995, 34: 187-192.
10. Stuart-Buttle CDG, Brown PJB, Price C, O'Neil MJ, Read JD. The Read Thesaurus - Creation and Beyond. In Pappas.C et al (Ed). *Medical Informatics Europe 1997*. IOS Press: 1997: 416-420
11. Stuart-Buttle CDG, Read JD, Sanderson HF, Sutton YM. A Language of Health in Action: Read Codes, Classifications and Groupings. In Cimino JJ (Ed). *Proceedings of the 1996 AMIA Annual Fall Symposium*. Philadelphia: Hanley & Belfus: 1996: 854-857.
12. Schulz EB, Barrett JW, Brown PJB, Price C. The Read Codes: Evolving a Clinical Vocabulary to Support the Electronic Patient Record. *Proceedings of Conference Toward an Electronic Health Record Europe*. Newton: CAEHR., 1996: 131-140.
13. Cimino JJ, Hripcsak G, Johnson SB, Clayton PD. Designing an Introspective, Multipurpose, Controlled Medical Vocabulary. In: Kingsland LC (Ed). *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care*. New York: IEEE Computer Society Press, 1989: 513-518.
14. Grishman R, Kitteredge R (eds). Analyzing Language in Restricted Domains: Sublanguage Description and Processing. Hillsdale, New Jersey: Erlbaum Associates, 1986.
15. Schulz EB, Barrett JW, Maclean AR, Brown PJB. A simple, maintainable system for identifying subsets of a modern controlled vocabulary. [accepted for publication]
16. Lindberg DAB, Humphreys BL, McCray AT. The Unified Nedical LAnguage System. Meth Inform Med, 1993; 32: 281-291.
17. Fuller JH. Mortality trends and causes of death in diabetic patients. Diabete et Metabolisme, 1993; 19: 96-99.
18. Sonksen P. Harris A. Jarman B. Perceived strategic and practical problems in the use of information technology for quality improvement in diabetes care in the United Kingdom. Diabete et Metabolisme, 1993; 19: 80-88.
19. Brown PJB, Cowley C. Realising the benefits of datasets for the care of diabetics in primary care. *Proceedings of the Annual Conference of the Primary Health Care Specialist Group 1997* [accepted for publication]]
20. Kalra D. Electronic health records: the European scene. BMJ 1994, 309: 1358-61.